1-1-2009

Folksonomy with practical taxonomy, a design for social metadata of the virtual museum of the Pacific

Peter W. Eklund
University of Wollongong, peklund@uow.edu.au

Peter Goodall
University of Wollongong, pgoodall@uow.edu.au

Timothy Wray
University of Wollongong, twray@uow.edu.au

Vinod Daniels
The Australian Museum

Melanie Van Olffen
The Australian Museum

Folksonomy with Practical Taxonomy, a Design for Social Metadata of the Virtual Museum of the Pacific[1]

Peter Eklund†, Peter J. Goodall†, Tim Wray†, Vinod Daniel‡ and Melanie Van Olffen‡

***Abstract*—The Virtual Museum of the Pacific is a Digital Ecosystem that engages members of several communities, each with their own ontological relationships with the Pacific Collection of the Australian Museum. The Virtual Museum of the Pacific is intended to support on-line community interaction using social-media technologies to extend the annotation of objects to suit the stakeholder's own needs. The success of the system depends on leveraging the diffusion of language and encouraging a conversation between on-line communities. In this paper we explore the relationships between stakeholders, folksonomy and taxonomy, to reveal the design forces on our digital ecosystem. Our analysis defines the scope for the social tagging component that progresses the design of our data model and gives us some confidence that we are capturing the right data for the system's development into the future.**

***Index Terms*—Information Technology in Literature and Art, Social Media, Taxonomy, Virtual Museum, Folksonomy, Ontology, Information System Design, Data Modelling, Access Control.**

I. Introduction

The Virtual Museum of the Pacific (VMP) is an environment for exploring and defining the relationships among a selection of the 60,000 objects in the Pacific Collection of the Australian Museum (AustMus or The Museum). The main motivation in the experiment of the VMP is to provide better access to the Museum's Pacific Collection for a wider variety of stakeholders and to give those communities a useful mechanism for accessing and annotating objects that are important to them.

The relationships among the objects of the collection are explored via a rich Internet client using web-services provided by our Formal Concept Engine. These services are used as input for generating web pages (Fig. 1) that assist user navigation by unobtrusively rendering a concept view with links to its upper (more general) and lower (more specialised) neighbours. The navigation paradigm is based on a technique called Formal Concept Analysis [1] and the design results from more than 10 years of research, development and testing.

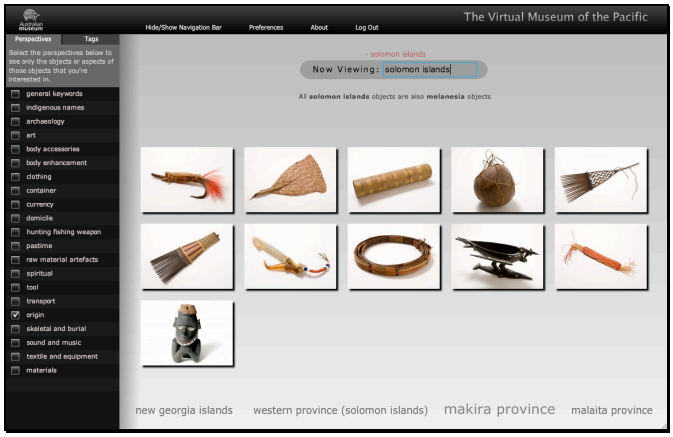


Fig. 1. Browsing a Formal Concept 'Solomon Islands'

Navigation of the pacific collection is dependent on relationships and attributes described in the metadata associated with the objects in the collection. This paper describes how community interaction through tagging, annotation, and metadata management influences the design of the Virtual Museum of the Pacific.

II. The Origins of Metadata within the VMP

The metadata used by the VMP for navigation and discovery within the collection is imported from the Australian Museum's Collection Management System (CMS). The current CMS is the third effort by the Museum to computerize its records of the Pacific collection.

To understand the evolution of the Pacific collection's metadata we give an overview of the typical life cycle of records. The Australian Museum (AustMus) acquired the objects in its Pacific Collection from many sources over the last 150 years. The process of adding an object to the collection is reasonably uniform and best illustrated by an example. The 'fish hook' (shown in Fig. 2) was entered into the AustMus 'Register of Ethnology' on September 22, 1971. This registry entry is the first association of collection metadata with the object, and instantiates its registration number. This is the initial source of the 'user warrant' [2] for the vocabulary associated with the object.

†School of Information Systems and Technology, The University of Wollongong, Northfields Avenue, NSW 2522, Australia, email peklund@uow.edu.au, twray@uow.edu.au, pjgoodall@gmail.com

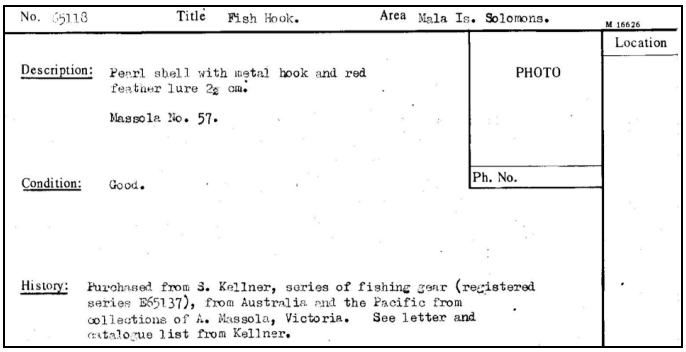
‡The Australian Museum, 6 College St, Sydney NSW 2000, Australia, email Vinod.Daniel@austmus.gov.au, Melanie.VanOlffen@austmus.gov.au

This work was supported by an Australian Research Council linkage grant LP0884075.

 ISBN: 978-981-08-3029-8

Fig. 2 - A fish hook from the Solomon Islands

By user warrant we mean that the staff that entered the object (the 'user') in the register have 'warrant' to generate its description details.



No. 65118 Title Fish Hook. Area Mala Is. Solomons. M 16626

Location

Description: Pearl shell with metal hook and red feather lure 2g cm.

Masgola No. 57.

PHOTO

Condition: Good.

Ph. No.

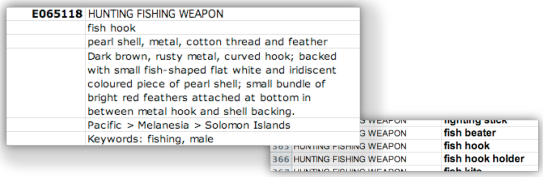
History: Purchased from S. Kellner, series of fishing gear (registered series E65137), from Australia and the Pacific from collections of A. Masgola, Victoria. See letter and catalogue list from Kellner.

Fig. 3 - Index card for the fish hook

At some later point in time an index card (see Fig. 3) was created which included the object's provenance, and more detailed descriptive text, and (on its reverse) the object's physical measurements. Later, as objects are added to the CMS, they are further described, and have a simple, practical corporate taxonomy applied to them. The spreadsheet documenting the Museum's taxonomy presents the 'organizational warrant' [2] for the metadata. The AustMus Archaeology and Anthropology taxonomy is two-level, with 27 categories and 709 object types distributed across those categories. The taxonomy provides a framework for describing objects in the collection and by organizational warrant we mean that it is 'warranted' or authorized within the 'organizational' context of the Museum.

From information collected during preparation of an initial 400 objects for the prototype of the VMP, we estimate that about 50 percent of the objects in the Pacific collection have an entry in the CMS, and nearly all objects need metadata cleaning to bring them up to a uniform high quality or exhibition standard. This involves normalizing spelling and thesaurus checking, for instance testing whether "mother of pearl" or "pearl shell' should be used or whether a "dagger" should be tagged as such or with a preferred term "knife". We estimate that an average of one hour's effort per object is required for basic metadata cleaning, and another hour to write an interpretive label (reminiscent of the descriptive card in a museum exhibition case). So, while the metadata adds enormously to the value of an object for research and Web-based exploration, there is a significant cost involved in establishing an adequate information base for it.

Fish Hook record 'E065118', as it appears within the existing CMS and spreadsheet classification scheme



E065118	HUNTING FISHING WEAPON
	fish hook
	pearl shell, metal, cotton thread and feather
	Dark brown, rusty metal, curved hook; backed with small fish-shaped flat white and iridiscent coloured piece of pearl shell; small bundle of bright red feathers attached at bottom in between metal hook and shell backing.
	Pacific > Melanesia > Solomon Islands
	Keywords: fishing, male

G WEAPON fish beater

366 HUNTING FISHING WEAPON fish hook

366 HUNTING FISHING WEAPON fish hook holder

Assumptions

An item has a distinct name eg. 'fish hook', that belongs in one category eg. 'hunting fishing weapon'.

An item's materials can be, to an extent, tokenised so that they can be expressed as a set of attributes and tags which are applied to itself or other objects.

An item's origin is inherently hierarchical - 'Solomon Islands' is located in 'Melanesia' which is located in 'The Pacific'

Keywords such as 'male' and 'fishing' are formally used to describe additional attributes of objects indicating their purpose. Incidentally, these keywords may also co-exist as item names belonging to a formal category.

Derived 'Tags', forming a subset of the formal taxonomy based on inferences within the object's metadata and AustMus thesaurus

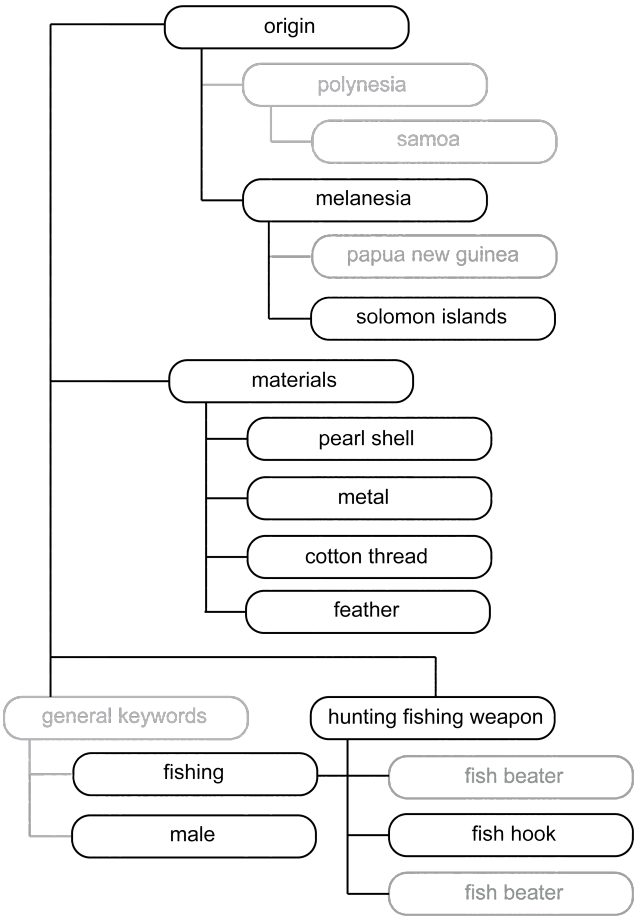


Fig. 4 - Inferred taxonomy from existing data model

In addition to the resources and time required to bring

metadata to exhibition standard, the formal taxonomy itself must be extracted from the collection so that a sufficient vocabulary of tags can be formed to fully describe objects within each of their facets and dimensions. As the key navigation features within the VMP are fully dependent on a reliable association between objects (for instance, the 'fish-hook' described in Figs. 2. & 3.) and their set of attributes, there was a significant challenge in translating existing data models and classification schemas from the Museum into a rich set of versatile tags that are atomic, multi-dimensional and hierarchical. Fig. 4 demonstrates this transformation process.

An ideal taxonomy derived from this process would cover multiple dimensions, describing not just the objects' common names (referred to as an 'item name' by the Museum), but also their descriptive terms such as materials, origin, and indigenous names, along with abstract terms related to their function or cultural significance.

The derived 'tags' – which can be used to describe an object in these dimensions – must be atomic and unambiguous. One of the difficulties with this process was that some reserved words used by the Museum had different meanings when applied in different contexts. For instance, the term 'clay' could be tokenized as a tag and used to refer to either the composition of the artefact or the whole artefact itself as a piece of clay. This use of tags as heteronyms creates problems in the assumptions and rules that were applied in translating and extracting 'tags' from the existing data models into a formal taxonomy as shown in Fig. 4.

The organization of tags into implicit categories and hierarchies allows for the partitioning of certain groups of tags (and their associative objects) into differing facets, which could then be explored by user communities. For instance, an anthropologist specializing in hunting and fishing weapons of the Pacific Islands can search those groups of associated objects by narrowing their search to that particular facet – which we call a 'perspective' within the VMP. Cross-relationships between perspectives can be exploited further – often revealing interesting or previously hidden findings by intersecting sets of objects inferred from common tags. For instance, the tag categories of 'origin' and 'hunting fishing weapon' can be enabled as perspectives, in which case the VMP reveals the relationship (if any) between certain types of hunting artefacts in certain locations.

Fig. 4 also demonstrates the use of data that is naturally hierarchical in character – 'origin' is an example of this, where the origin of an artefact can be drilled down to its area group, country, state or individual island. The VMP provides a visual metaphor for drilling-down into specific hierarchies, by visually growing and shrinking the groups of objects as the search terms (the set of tags) become more general or more specific respectively. This hierarchy can be combined with other semantic dimensions to produce an effective way of navigating thematically similar objects by inferring cross-dimensional relationships along with the ability to infer sub- or super-sets of objects.

The translation of the Australian Museum's existing metadata into a formal taxonomy - which is then applied into the VMP - presents an interesting and novel application of the collection that has sparked considerable stakeholder interest in communities that wish to explore and annotate the objects. A further discussion will follow on the different user communities that will interact with the collection, along with issues concerning the management of a user-driven bottom-up folksonomy and its compatibility with the derived taxonomy discussed above.

III. Kinds of Communities using the Virtual Museum

There are many types of communities which we expect to use the VMP for object discovery and annotation. Each community may create its own specific annotations, and may be influenced by the annotations of other communities.

There are several evident stakeholder groups that can be inferred as intersecting communities – indigenous populations, museum staff, independent researchers, students, hobbyists and anthropologists are a non-exhaustive list. The first of these is the original Pacific Island community from which the object was collected or acquired. They could be said to have had, or still have, a thorough understanding of the object's cultural significance and practical use.

There are then possibly multiple transactions between people 'collectors' who have possession of the object before it reaches the Museum. The documented information passed on by collectors along with classification and provenance documents form the foundation of the metadata associated with the object at the Museum. Given that the Pacific Collection has existed for more than 150 years and that the original Pacific Island communities may have transformed considerably over that time, the information about an object in the possession of the Museum becomes increasingly important in defining its meaning.

The minimal set of communities likely to make use of the VMP will be scholars, the diaspora from the originating communities, the communities in their original homeland and the general public.

Each of these communities attaches different subjective significance and vocabulary to the objects. The interactions and overlaps between private and public views of the objects, and the community's opportunity to leverage one another's knowledge in a respectful way is a source of enthusiasm for many involved in the Virtual Museum of the Pacific.

IV. Critical Mass, Tagging Intensity, Community Size and Involvement

A collection of 60,000 objects from any source without metadata is a daunting prospect for exploration. Imagine a library of books with blank covers, and no cataloguing or ordering of books on its shelves. For this reason, the existing metadata provided by the Museum is of extraordinary value. Without it, every object would be 'lost'. The VMP uses metadata from the Museum's CMS to seed the relationships among the objects. Once this is in place, the communities have the opportunity to find the objects most important to them.

If each of the communities has access to tools to tag, annotate and re-focus the visible vocabulary around objects that they have found interesting, they are able to adjust the conversation to improve the relevance to themselves, as well as improve, correct and extend the quality of the metadata.

The effort each community makes in object annotation can, where appropriate, influence the conversation about objects in other communities. For instance, it is almost certain that if the original Pacific Island community makes public additions to descriptions of objects, that these changes in vocabulary will affect the language and taxonomy used by scholars at the Museum, improving the coverage and timeliness of categorization and other annotation [4,3]. Access to objects that may have few examples in their homeland will also encourage discussion and knowledge in their communities of origin.

Each of the likely communities will have differing profiles for tagging. Therefore, the ontology of the VMP will be a system of interacting communities and their annotations.

Current terminology in the literature speaks of broad and narrow folksonomies. 'narrow' commonly describes a user tagging resources for their own purposes, and 'broad' usually referring to collaborative tagging by a large number of users intent on knowledge sharing [5,6]. We believe that in a system of communities, each more or less distinct from the others, it will become more appropriate to evolve 'breadth' as a qualifier for a folksonomy. 'breadth' may come to represent the size of the community and the rate of diffusion of its vocabulary within other communities.

It is likely that communities using the VMP will substantially vary in size and activity. There is a reasonable hope that useful semantics will emerge from the activities of communities of all sizes [7,8].

The VMP is a practical experiment, directed towards producing a useful environment for constructive social engagement with the Pacific Collection. Because of this it is important that we capture the data fundamental to enabling rich toolsets for community engagement. The most basic data required for analysis is a core triple of *<user, resource, {tags}>*, augmented by a timestamp; this represents a 'post' event [8]. The other important association to capture is that between the user and any groups they are members of. This user and group association will help considerably separating the semantics emerging from each group, and reduce the apparent 'noise' that would occur if many small groups activities were aggregated as one large tag-space.

Additionally, the partitioning of users into groups can introduce a level of control with the quality of both the tagging of objects (the associations between the tag and the object) and the definition of customized tag groups – or 'perspectives' – discussed in Section II. This is to ensure protection of the object tags and their taxonomies from abuse from nuisance tagging.

A user may be a member of one of more groups, in which each group has a certain level of permission. Some groups may or may not be able to tag objects, whereas other groups may or may not be able to create their own tags or folksonomies or interact with an existing folksonomy. This control is crucial as the clarity and multi-dimensionality of the tag hierarchy is a key determinant in providing cross-dimensional relationships or interpretations of museum objects, especially as the interpretive description of an object can be highly influenced by its context and user community.

The level of involvement that a group would be allowed to have with the collection and folksonomy is dependent on the relevance, interest and sentimental value of the objects as determined by a group's administrator. Naturally, indigenous communities and curators would have a high level of permission and access whereas unregistered users or the general public will have more restricted access.

V. Access Control Model

The design of the access control model of the VMP is carefully considered in order to achieve a balance between accommodating the interests of the user community and preserving the integrity of the formal taxonomy derived in Section II. Additionally, restricted access to groups such as the general public or casual users of the VMP are critical requirements in consideration of intellectual property or other sensitive issues concerning the exposure of the artefacts to broader communities.

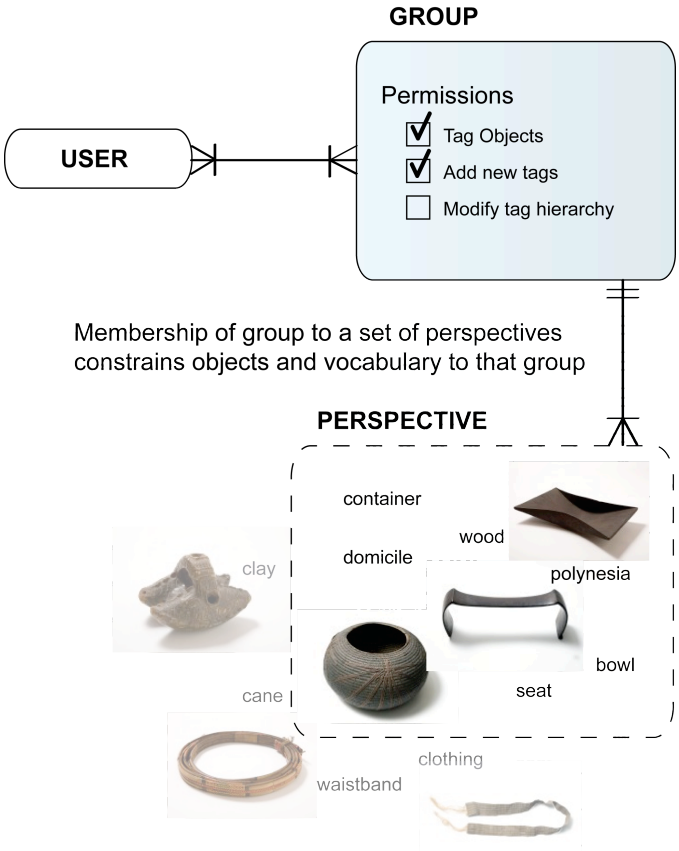


Fig. 5 – Access control model describing relationship between users, groups and perspectives

As discussed in Section IV, the roles and permissions of registered users within the VMP are primarily determined by

their group membership. Groups can be either public or private, where users can opt-in to join a group or be registered exclusively by invitation only. Fig. 6 identifies several user groups with varying levels of permission in terms of their ability to view, edit and delete objects, tag hierarchies and perspectives. This table represents a sub-set of permissions made available to the users, and represents the first dimension of access control, which is that of a role-based one. Note that although four pre-determined roles appear to exist, they can be customized according to the permissions set by the administrators of that group.

	Collection Managers and Curators	Research Specialists	Indigenous Communities	General Public (Registered Accounts)
Viewing Object Metadata				
Label / Description	X	X	X	X
Category / Item Name	X	X	X	X
Acquired Date	X	X	X	X
Materials	X	X	X	X
Location	X	X	X	X
Dimensions	X	X	X	X
Registration Number	X	X	X	
Keywords and 'tagged' attributes	X	X	X	X
Indigenous Names	X	X	X	X
Managing and Editing Objects				
Add a tag to an object	X	X	X	
Remove tag from an object	X		X	
Add a new object	X		X	
Change / Upload an object's image	X	X	X	
Edit object metadata (excl. tags)	X		X	
Managing and Editing Attributes / Tags				
Define a new attribute / tag	X		X	X
Change the definition of an attribute	X		X	X
Remove an attribute / tag	X		X	X
Managing and Editing Perspectives				
Define a new perspective	X	X	X	X
Assign / Unassign attributes to perspective	X	X	X	X
Remove perspective	X	X		
Wikis and Rich Multimedia Content				
View Wikis assigned to objects	X	X	X	X
Commit changes to Wikis	X	X	X	
Attach multimedia or files to objects / wikis	X	X	X	

Fig. 6 – Preliminary access control table

The second dimension relates to the restricted set of objects a group is allowed to interact with along with a restricted vocabulary set – known as a perspective, that assigns semantic meaning to those objects as shown in Fig. 5. For instance, an indigenous group from a certain region of the Pacific may have a high set of permissions relating to the ability to extend the vocabulary of tags (as they can provide meaningful indigenous descriptions of those objects beyond the means of the museum curators or general public) – but their perspective may be limited to a particular sub-set of objects from that certain region. This model can be extended to other user groups where restricted permissions, objects and vocabulary sets are necessary in order to alleviate concerns surrounding the exposure of objects to the general public and abuse of the formal taxonomy.

The clustering of users into groups, which are then defined by permissions and perspectives, allows for inter- and intra -group collaborative efforts to be encouraged while still retaining control over the exploration and tagging of objects within the Pacific collection. This model ensures vibrant community participation and folksonomy generation with little or no risk to the valuable data contained within the researched metadata and extracted formal taxonomy.

VI. Taxonomy and Annotation for the Long-haul

The Australian Museum's Pacific Collection is already older than any person, and is intended to be perpetual. How do annotations behave over a long period of time? Terminology in any community changes as understandings evolve; nomenclature drifts with time and contemporary tagging frequency changes. Historical tags compete with current usage for our attention.

It is possible that what was once a relevant taxonomy in a subject area which had a high currency could be made less relevant by a contemporary less frequently used taxonomy. Technical subjects discounted over time are good examples – for instance the concepts of 'phrenology' and 'phlogiston' were both popular in their time but are now defunct.

Likewise, terms used to describe, classify or evoke the functions or cultural significance of artefacts may change over time, and hence the classification models or terminology may adapt as such. Given that a user group has enough privileges to do so, they would be able to define or re-define a classification schema to suit contemporary trends.

VII. Interacting Folksonomies and Taxonomies

The Australian Museum, after considerable experience and practice, chose to create and administer their own corporate taxonomy. This formally managed taxonomy with its own descriptive vocabulary is embodied in the Museum's CMS. While the annotations and tags applied by stakeholder communities to objects in the collection are likely to be less formal and of the type often referred to as a 'folksonomy' [9], it is our contention that the warrant of all formal taxonomies emerges from the vocabulary of some interest community. We expect to facilitate the emergence of community derived, dynamic taxonomies from the social media that the VMP will support, as well as contributing to the evolution and relevance of the formal taxonomies of museums.

We believe that the digital ecosystem of interacting

communities we expect the VMP to become, these terms describe the endpoints but obscure the probability of there being a continuum of formality and breadth. We expect that the interaction between formal taxonomies and the communities' folksonomies will enrich both, keeping the former fresh, and up-to-date, and provide some stability and common vocabulary for the latter, creating a useful metadata digital ecology [10,11].

The ontology and the communities that create them, associated with long-lived collections like those of the Australian Museum, evolve over time. The museum had a particular taxonomy 100 years ago, and another 50, 15,10 and 5 years ago. What influenced the changes in this taxonomy? Some examples are:

- Changing culture and understanding of the subject domain.
- Evolving interaction with indigenous communities.
- Clash of technology – an attempt to apply 'big' taxonomies to the collection, which was later rejected.

VIII. Keeping Tags relevant and Concise

Formal taxa are inevitably influenced by community usage [7,8]. In the context of the VMP, folksonomy represents a readily available representation of community usage that can be readily analysed.

Much of the discussion of folksonomies mentions the occurrence of typographic errors when applying tags, but does not suggest the use of stemming, thesauri or other Information Storage and Retrieval tools to help manage the intrusion of errors. Applying algorithms to the tags after posting by a user may introduce misinterpretations, thus is it seems more useful to provide support and suggestions from tools before the user commits the post, thus ensuring the user's intent is captured more accurately [3].

Some of these tools can include data validation to determine if a new tag already exists, the use of edit distance or other string-based metrics to compare new tags with existing ones within the folksonomy or taxonomy, and visual tools for graphically navigating and modifying tag hierarchies to ensure that the tag is placed within it's relevant category or perspective if it is being added to a formal taxonomy.

IX. Other Aspects of Social Media and the Pacific Collection

The Australian Museum's Pacific Collection contains items of cultural significance. This implies a challenge to minimize possible offense given and taken through visibility of commentary by individuals and communities. It is not hard to see that any management of visibility of the annotation of objects in the collection will affect the evolution of descriptions of those objects by individuals and communities.

There are responsibilities of management that need to be provided to control impolite posting and graffiti [11]. This is a subject of further study by us, but not covered within this paper.

X. Conclusions

The Virtual Museum of the Pacific is a digital ecosystem that allows social tagging by its stakeholders. In this paper we have described the folksonomy literature and its relationship to the Australian Museum's Pacific taxonomy. We have presented an access control model that describes how social media resulting from community tagging will be captured and treated. Our conclusions is that the formal taxa be maintained separately from the folksonomy tags and our access control use cases give us confidence that our design of the Virtual Museum of the Pacific will meet the requirements of stakeholder communities.

References

[1] P. Eklund and P. Goodall, Tim Wray, B. Bunt, A. Lawson, L. Christidis, V. Daniels, M. Van Olffen, "Designing the Digital Ecosystem of the Virtual Museum of the Pacific", 3rd IEEE International Conference on Digital Ecosystems and Technologies, IEEE Press, 2009.

[2] E.G. Fayen, "Guidelines for the construction, format, and management of monolingual controlled vocabularies: A revision of ANSI/NISO Z39. " http://www.techstreet.com/cgi-bin/detail?product_id=1262086, 2005

[3] S. Hayman, "Taxonomy Directed Folksonomies," *New Developments in Social Bookmarking, Ark Group Conference: Developing and Improving Classification Schemes, Sydney June*, 2007, http://www.ifla.org.sg/IV/ifla73/papers/157-Hayman_Lothian-en.pdf

[4] J. Voss, "Tagging, Folksonomy & Co-Renaissance of Manual Indexing?," *Proceedings of the International Symposium of Information Science* 2007.

[5] T. Vander Wal, "Explaining and showing broad and narrow folksonomies," *Vanderwal. net, http://www.vanderwal. net/random/entrysel.php*.

[6] T. Vander Wal, "Folksonomy Explanations :: Off the Top :: vanderwal.net."

[7] M. Lux, M. Granitzer, and R. Kern, "Aspects of broad folksonomies," *Database and Expert Systems Applications, 2007. DEXA'07. 18th International Conference on*, pp. 283-287, 2007.

[8] C. Cattuto, V. Loreto, and L. Pietronero, "Semiotic dynamics and collaborative tagging," *Proceedings of the National Academy of Sciences*,. **104**, 2007, p. 1461.

[9] Vander Wal, "Folksonomy :: vanderwal.net," 2007.

[10] L. Rosenfeld, "Bloug: Folksonomies? How about Metadata Ecologies?" Jan. 2005.

[11] D. Barbosa, "Taxonomy Folksonomy Cookbook," 2008.